Slicing

Received: 8 May 2019 | Accepted: 21 December 2019

DOI: 10.1111/2041-210X.13352

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Slicing: A sustainable approach to structuring samples for analysis in long-term studies

Sil H. J. van Lieshout[1] | Hannah Froy[2,3] | Julia Schroeder[4] | Terry Burke[5] | Mirre J. P. Simons[5,6] | Hannah L. Dugdale[1]

[1]School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds, UK; [2]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK; [3]Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway; [4]Department of Life Sciences, Imperial College London, Ascot, UK; [5]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK and [6]The Bateson Centre, University of Sheffield, Sheffield, UK

Correspondence
Sil H. J. van Lieshout
Email: sil.vanlieshout@gmail.com

Funding information
Leeds Anniversary Research Scholarship; Sir Henry Wellcome Fellowship, Grant/Award Number: WT107400MA; University of Sheffield Vice-Chancellor's Fellowship; Natural Environment Research Council, Grant/Award Number: NE/J024597/1, NE/N013832/1 and NE/P011284/1

Handling Editor: Laura Graham

Abstract

1. The longitudinal study of populations is a core tool for understanding ecological and evolutionary processes. Long-term studies typically collect samples repeatedly over individual lifetimes and across generations. These samples are then analysed in batches (e.g. qPCR plates) and clusters (i.e. group of batches) over time in the laboratory. However, these analyses are constrained by cross-classified data structures introduced biologically or through experimental design. The separation of biological variation from the confounding among-batch and among-cluster variation is crucial, yet often ignored.
2. The commonly used approaches to structuring samples for analysis, sequential and randomization, generate bias due to the non-independence between time of collection and the batch and cluster they are analysed in. We propose a new sample structuring strategy, called slicing, designed to separate confounding among-batch and among-cluster variation from biological variation. Through simulations, we tested the statistical power and precision to detect within-individual, between-individual, year and cohort effects of this novel approach.
3. Our slicing approach, whereby recently and previously collected samples are sequentially analysed in clusters together, enables the statistical separation of collection time and cluster effects by bridging clusters together, for which we provide a case study. Our simulations show, with reasonable slicing width and angle, similar precision and similar or greater statistical power to detect year, cohort, within- and between-individual effects when samples are sliced across batches, compared with strategies that aggregate longitudinal samples or use randomized allocation.

Mirre J. P. Simons and Hannah L. Dugdale contributed equally to this study.

[Correction added on 4 March, after first online publication: Supporting information file name has been changed from 'mee313352-sup-0003-DataS2.R' to 'mee313352-sup-0003-DataS1-Year.R']

4. While the best approach to analysing long-term datasets depends on the structure of the data and questions of interest, it is vital to account for confounding among-cluster and batch variation. Our slicing approach is simple to apply and creates the necessary statistical independence of batch and cluster from environmental or biological variables of interest. Crucially, it allows sequential analysis of samples and flexible inclusion of current data in later analyses without completely confounding the analysis. Our approach maximizes the scientific value of every sample, as each will optimally contribute to unbiased statistical inference from the data. Slicing thereby maximizes the power of growing biobanks to address important ecological, epidemiological and evolutionary questions.

KEYWORDS

ageing, biobank, cross-classified, long-term studies, mixed models, nested, slicing, telomeres

1 | INTRODUCTION

Individuals and populations are shaped by ecological and evolutionary processes which generally occur over many years or decades (Clutton-Brock & Sheldon, 2010). Consequently, long-term studies are key in determining the proximate and ultimate causes of biological processes. Sampling a population repeatedly over individual lifetimes and across multiple generations allows quantification and separation of genetic variation from environmental variation and estimation of such effects with appropriate precision and statistical power (Martin, Nussey, Wilson, & Reale, 2011; van de Pol, 2012). However, statistical analyses of such comprehensive biological datasets are often complex due to hierarchically structured data and difficulties in separating variation from sources of interest and confounding variables.

Due to the hierarchical nature of biology, for example, phenotypic traits nested within individuals, individuals nested within social groups and social groups nested within populations (Figure 1a), appropriate statistical methods are required that model the hierarchical structure of biological datasets. While nested designs, either natural or through experimental design (Figure 1a), can be analysed in linear models, this inflates the degrees of freedom and thus reduces statistical power (Gelman, 2005; Quinn & Keough, 2002; Underwood, 1997). A better approach is the mixed model framework, which estimates fixed effects while flexibly accounting for the variance explained by random effects, incorporating multilevel hierarchies in data (Bolker et al., 2009; Gelman & Hill, 2006; Snijders & Bosker, 2011; Zuur, Ieno, & Elphick, 2010). However, in cross-classified designs (Table 1), where one individual is associated with more than one batch (Figure 1b) or even more than one cluster (Figure 1c), advanced statistical methods to estimate fixed effects and variance components are required compared to nested designs (Schielzeth & Nakagawa, 2013). While cross-classified data structures in short-term studies are often the result of the experimental design (e.g. cross-fostering), in long-term studies the timing of the analyses of data often naturally leads to cross-classification of data (Figure 1b,c).

In long-term studies, the individual-based collection of longitudinal data and biological samples from natural or laboratory populations produces large, continuously growing biobanks (Clutton-Brock & Sheldon, 2010). Through laboratory analyses, these biobanks provide information on, for example, individual telomere length (Boonekamp, Mulder, Salomons, Dijkstra, & Verhulst, 2014; Fairlie et al., 2016), serological values (Andraud, Casas, Pavio, & Rose, 2014; Telfer et al., 2008) and genetic variation (Berry, England, Marriott, Burridge, & Newman, 2012; Tollenaere et al., 2012). However, the laboratory analysis of samples from growing biobanks is often conducted on separate groups of samples over time (e.g. after each fieldwork season, each year or coinciding with grant cycles). Such a group of samples—a cluster—will be collectively analysed under similar conditions, but these conditions might differ between clusters (e.g. different analyst, machine or month). Samples within a cluster are often further subdivided into batches (e.g. qPCR-plates) where, again, samples are analysed under similar conditions, but conditions may vary between batches (e.g. different reagents or day). While batches are nested within clusters, the continuous collection of samples in the field and intervals between laboratory analyses result in longitudinal samples from a single individual that may not be nested within batches or even clusters, causing cross-classified data structures in long-term studies (Gelman & Hill, 2006; Figure 1b,c).

Cross-classification of data induces variation that can be confounded with the independent variables of interest, which can reduce the ability to compare results across samples and draw reliable conclusions (Greenland, Robins, & Pearl, 1999; Schielzeth & Nakagawa, 2013). This is problematic if cross-classification is not explicitly accounted for, or there is not sufficient cross-classification to disentangle these sources of variation with high statistical power. For example, temporal variation or, where multiple populations are studied, spatial differences in resource availability can be confounded with laboratory analysis when samples are analysed after each period of collection, resulting in a failure to separate the effects of resource availability and laboratory analysis on a response variable. The experimental design and therefore the method in

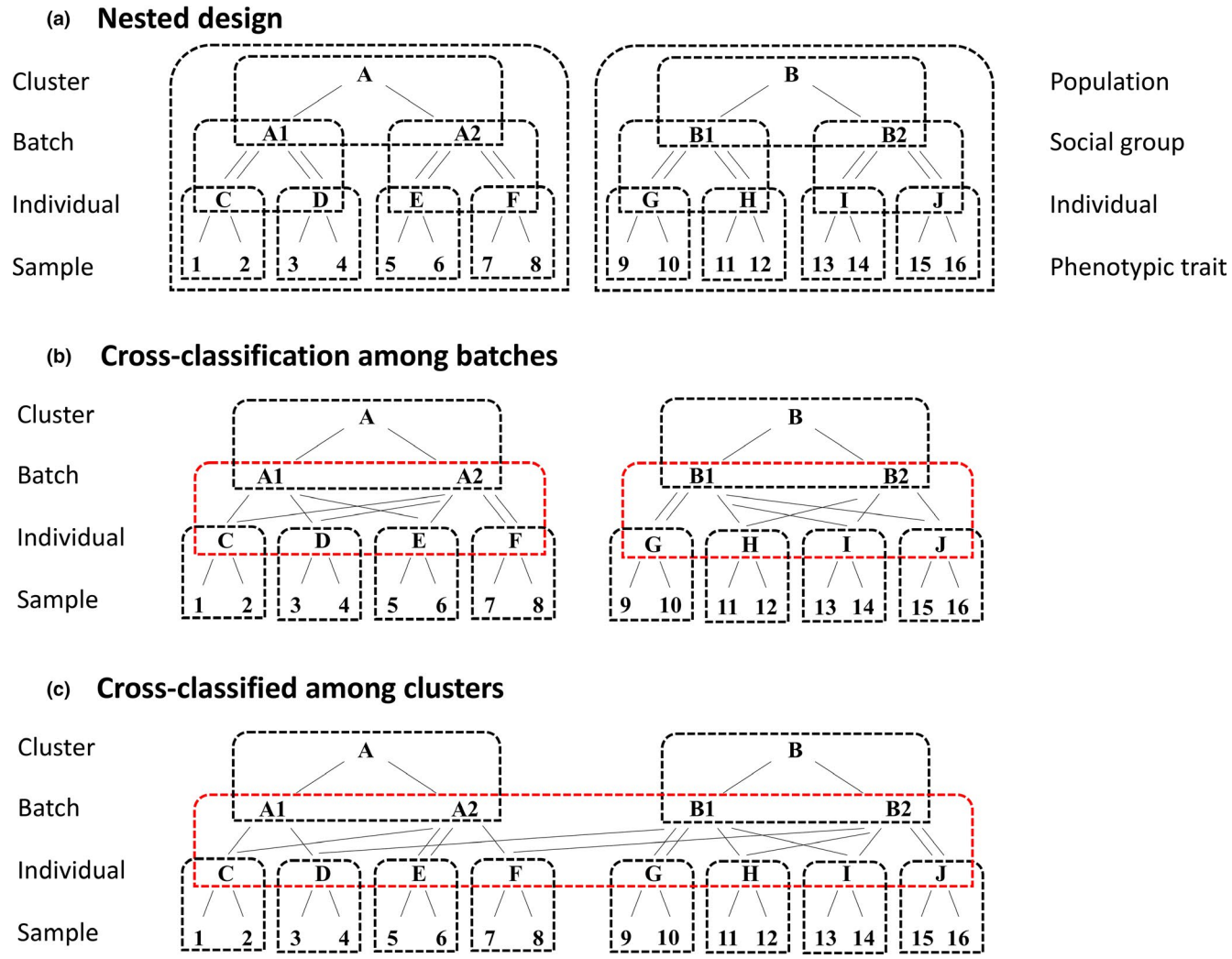

FIGURE 1 Schematic of nested and cross-classified data structures: (a) with a nested design applied to laboratory analyses (left) and populations (right); (b) cross-classification of data among batches that is confounded by time of analysis; and (c) cross-classification common in longitudinal data in laboratory analyses across clusters. Black dashed delineation indicates nested, whereas red dashed delineation indicates cross-classified structures

TABLE 1 Definitions of key terms

Term	Explanation
Batch	A set of analysed samples that are inherently dependent on one another, e.g. all using the same standard curve, machine, time of day, technician or that are equally affected by any other source of variation
Cluster	A set of samples that are distinct in the timing of their analysis; this typically includes multiple batches nested within clusters
Nested	All units at a lower hierarchical level are within one unit at a higher hierarchical level
Cross-classified	Units at a lower hierarchical are associated with more than one unit at a higher hierarchical level
Within-individual effects	Longitudinal changes within individuals in a repeatedly measured trait
Between-individual effects	Comparison of differences in mean traits among individuals
Statistical power	The ability to reject the null hypothesis when false, quantified by the proportion of significant values ($p < .05$) out of the total
Precision	The degree to which simulations provide similar results, quantified by the absolute difference between the 75% and 25% percentile divided by the median
Slicing	Recently added samples analysed in clusters together with previously obtained samples, ensuring statistical independence of collection time and cluster

structuring samples for clusters and allocating samples to batches requires consideration to cope with cross-classified data structures and confounding variables.

While relatively few studies report the approach used to structure samples into clusters, currently two main approaches are used, and both are prone to confounding effects and cross-classified data structures. First, sequential structuring of samples to clusters: analysing samples in clusters, in the same order in which they were collected (e.g. by year). This approach may be used, for example, in physiological studies (e.g. Takizawa et al., 2004) and has the advantage that samples can be analysed immediately without any issues in placing or labelling of samples. However, sequential structuring of samples confounds cluster with organizing variable (e.g. year) effects (Figure 2a). The second approach, randomization of samples from multiple years within a cluster, ensures that samples are sufficiently mixed to avoid confounds, and should already be standard practice (Figure 2b). The use of randomization is widespread in, for example, telomere length (e.g. Spurgin et al., 2017), disease (e.g. Swanson et al., 2015) and hormone analyses (e.g. Dantzer et al., 2013). However, this randomization approach requires a delay before analyses can be completed so that samples collected at different time points can be analysed together, and organizing variable and cluster effects can be separated. Furthermore, the randomization of large numbers of samples is time-consuming and detailed reordering of samples from the biobank is prone to error due to sample labelling and placing. Most importantly, however, is that after applying this randomization approach once in a long-term study, any subsequently collected samples cannot be directly compared to the previously randomized samples as they will be subject to statistically inseparable variation due to clustering of the samples already analysed. For example, randomizing two time periods of 4 years of sampling separately into two clusters results in uncontrollable variation between these two clusters and confounds the first 4 years in cluster one with the subsequent years in cluster two (Figure 2b), leading to cross-classified data structures (Figure 1c). Analysing the same samples multiple times in subsequent clusters can avoid this issue, often referred to as 'golden' or 'reference' samples. However, the additional costs or potential depletion of the 'golden' sample can make this approach difficult. More importantly, it is unclear how effectively one golden sample can control for among-batch and among-cluster variation. For example, the 'golden' sample might not be representative of all

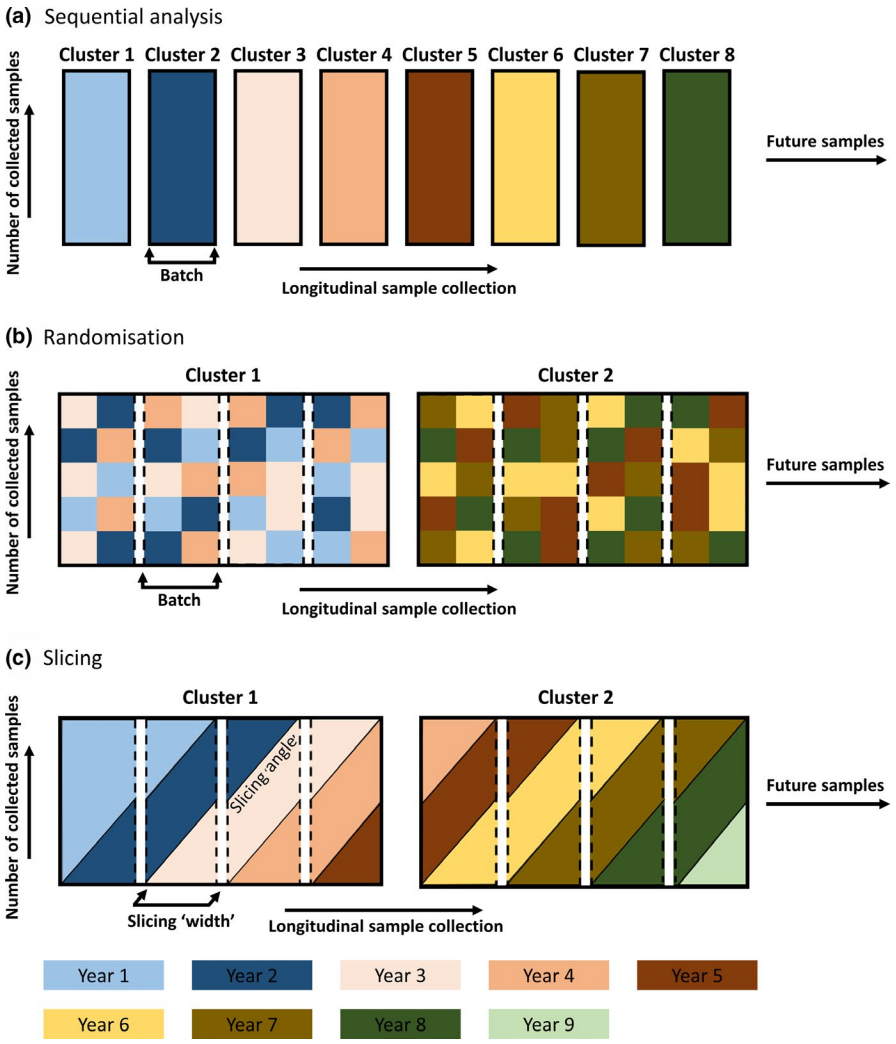

FIGURE 2 Schematic of three strategies to structure samples from the biobank. The sequential analysis strategy (a) can confound cluster and year, while randomization of multiple years within a cluster (b) prevents this confound but generates uncontrollable variation between clusters. The slicing approach (c) combines the advantages of these approaches and can be used to sequentially analyse growing biobanks while maintaining independence between cluster and associated variables. The biobank is sliced (e.g. by year), thereby analysing a set of continuously collected samples sequentially in each subsequent cluster. Each sample only needs to be analysed once, where different samples from the same slice are analysed across batches and clusters (e.g. years 4 and 5), which enables controlling for batch and cluster effects. Slicing width (frequency of new samples collected) and angle (degree of independence between slices) determine the level of statistical independence between clusters

samples, and the sample can degrade over time thus not returning the same value in different analyses. In short, these two popular approaches to structuring cross-classified samples do not fully account for among-cluster and among-batch variation, leaving an unknown amount of variance unquantified and thus compromising conclusions drawn from such studies.

The analyses of longitudinal data can be turned into a nested design when samples from a single individual are aggregated within a batch and cluster (Figure 1a). This is thought to increase the statistical power to detect within-individual effects. The reasoning is that longitudinal samples are then exposed to the same technical noise, which allows greater statistical power to dissect out the biology from batch effects (Beirne, Delahay, Hares, & Young, 2014; Herborn et al., 2014; Nettle et al., 2015; Pauliny, Devlin, Johnsson, & Blomqvist, 2015; Rius-Ottenheim et al., 2012; Sudyka et al., 2014). Although the aggregation approach may be optimal for certain questions and data structures (e.g. experimental studies where the focus is on within-individual changes, while having controlled for many other sources of variation), the increasing application of physiological assays in long-term studies requires a different approach because the aggregation of longitudinal samples in a single batch has four disadvantages. First, analyses need to be postponed until all samples from a single individual have been collected. Second, aggregation requires detailed picking and reordering of samples, which increases the likelihood of human error, sample mix-ups and therefore false conclusions. Third, confounding variables that covary with the individual samples taken from one individual are not effectively separated from batch or cluster (e.g. seasonal effects). Fourth, it is rare for within-individual variation in a trait to be the sole interest, often between-individual variation is of interest too, and aggregating individuals within a batch could reduce the ability to estimate between-individual variation when individuals are aggregated in and thus confounded with batch effects. Thus although assumed, it remains to be determined whether aggregation increases statistical power to detect within-individual effects so substantially that it would outweigh these four disadvantages. Hence, the approaches to structure samples for analyses in long-term studies suffer from confounding effects, cross-classified data structures and increased likelihood for human error, and cannot provide the comparable analyses of samples over time required in long-term studies.

Here, we present an approach to the analysis of samples from growing biobanks that, while maintaining statistical independence, accounts for among-cluster variation and controls for other potentially confounding effects (Figure 2c). Additionally, we provide a case study of this novel approach and subsequently test the assumption that aggregating longitudinal samples within batches results in greater statistical power to detect within-individual effects. We then discuss the analysis of long-term data and highlight the importance of statistical mixed models. While we will mainly consider the field of evolutionary biology, using telomere dynamics as an illustrative example, these considerations and techniques can be applied to a range of fields, including epidemiology, ecology and laboratory-based science.

2 | MATERIALS AND METHODS

2.1 | Slicing approach

We have developed a slicing approach to structure samples from growing biobanks, such that recently collected samples are analysed in clusters together with previously obtained samples, ensuring statistical independence of collection time and cluster. This approach can overcome the experimental design and statistical issues with cross-classification and confounding variables in long-term studies (Gelman & Hill, 2006; Greenland et al., 1999; Schielzeth & Nakagawa, 2013), by bridging batches and clusters together. This allows for a structured and a priori separation of the variation of interest from confounding variables and when combined with mixed models copes well with cross-classified data structures. The biobank is divided into slices (Figure 2c), where a slice reflects a group of collectively gathered samples (e.g. in the same year) analysed together. Samples from a slice can be sequentially allocated to batches and only need to be analysed once, with the benefit of needing less sample volume and less degradation of samples. Separate samples from the same slice can be analysed in different batches or clusters (Figure 2c), bridging batches and clusters together. Slicing uses a varying proportion of samples from each given sampling period (i.e. slices), sequentially analysed in a single cluster, to statistically account for temporal and cluster variation. Slicing therefore allows convenient sequential analysis while maintaining statistical independence.

Depending on the frequency at which new samples are obtained, the 'width' of the slices can be changed (Figure 2c). For example, low analysis frequency requires wider slices to account for among-cluster variation. This decision is directly related to the slicing 'angle' (Figure 2c), which determines the degree of independence of sampling year from cluster. For example, if there are environmental effects related to the collection time of samples, slicing samples by collection time (i.e. lower angle) removes possible confounds with cluster effects. For slicing to be effective across clusters, it requires multiple years/cohorts to be present within a single cluster and at least one of those years/cohorts to be present in a different cluster, allowing statistical separation of among-cluster variation and confounding effects. Depending on slicing width and angle, a slice should cover approximately one-third of each batch, when slicing across three batches, with three separate slices covering the same batch (Figure 2c and see Section 2.3). Such a strategy also naturally allocates samples of certain slices to batches in subsequent clusters, bridging clusters together (Figure 2c) and allowing control of among-cluster variation. As a general rule, when you have more confounding effects smaller slices (i.e. lower angle and smaller width) are required to be able to partition these confounding effects. Smaller slices lead to a greater statistical power

to separate potentially confounding effects within and between batches (as there are more slices within a batch and each slice occurs in more batches). Setting the slicing angle and width is a trade-off between statistical independence (assessing statistical power in the case of confounding effects) and the number of samples that remain unanalysed until the addition of newly collected samples. This latter point is a constraint, as the number of samples that can be analysed simultaneously will be reduced, if only slightly, by this approach. We argue that the creation of statistical independence and accounting for among-cluster variation are merits that outweigh this limitation.

2.2 | A case study: structuring samples for telomere length analysis in wild house sparrows

We provide a case study of how slicing can be applied to structure samples for analysis in a long-term (>20 years) study on a natural population of house sparrows (*Passer domesticus*) on Lundy Island, UK (Schroeder, Nakagawa, Rees, Mannarelli, & Burke, 2015). House sparrows are a relatively long-lived species (on Lundy: mean lifespan is 3.5 ± 1.4 *SE*, maximum lifespan is 9 years; Schroeder, Burke, Mannarelli, Dawson, & Nakagawa, 2012). The Lundy population has been systematically studied since 2000 and the adult population size varies between years (Simons, Winney, Nakagawa, Burke, & Schroeder, 2015). Immigration to and emigration from the island is low (0.5% of recruits; Schroeder et al., 2015), with an annual resighting probability of 0.91–0.96 (Simons et al., 2015). This closed island population on Lundy thus provides precise ages and life-history data for all individuals.

We use a subset of the Lundy dataset containing 12 years of data (2000–2011; Table S1), where the population consisted on average of 130 individuals that were blood sampled on average twice a year. The total biobank we selected for in this case study contains 2,733 samples from 515 individuals. The hypothesis to be tested is that telomere length and age are negatively associated within individuals, and therefore we will analyse all samples collected every 6 years (i.e. 12/6 = 2 clusters) with 12 qPCR plates (i.e. batches) in each cluster (Figure S1). Samples are analysed sequentially, where each sample is analysed once. A key consideration is to separate variation in sources of interest from confounding variables by analysing samples with different confounding effects in the same batch. This ensures that confounding effects (e.g. sampling year) are not fully confounded with attributes of batch. Slicing, where samples are sliced across batches within a cluster, can achieve such separation in combination with mixed models to statistically correct for known confounding effects (e.g. qPCR-plate).

In the Lundy sparrow example, we first determine the slicing width, which depends on the analysis frequency and number of samples collected in each year. The analysis frequency (i.e. 2 clusters) is relatively low which results in many sampling years within a cluster. The contribution of a confounding sampling year effect can be determined by comparison of within-year to between-year effects, which requires sufficient samples from a single year analysed in the same batch and therefore wider slices. However, the number of samples collected in each year varies markedly, resulting in a variable slicing width per year depending on the number of samples in each year (Figure S1). Second, we determine the slicing angle. Since the population density varied strongly between years, the slicing angle should be low (Figure S1). This way a single year crosses more batches which allows confounding effects (i.e. population density and year) to be separated from variation in sources of interest. Third, since the number of samples exceeds the preferred slicing width and angle, multiple batches with the same lay-out will be used (Figure S1). These slicing parameters result in at least three slices within a batch to enable the separation of confounding environmental effects (e.g. population density, sampling year) from laboratory effects (e.g. batch), when using mixed models (Gelman & Hill, 2006).

The slicing approach allows an accurate estimation of the relationship between telomere length and age. Since the Lundy sparrow study is ongoing, the slicing approach can be continued into new clusters without inducing new confounding effects. For comparison, the sequential approach would confound sampling year with batch effects while randomization of samples could result in human and technical errors. Additionally, randomization would not allow comparable analyses among clusters or flexible inclusion of current data in future analyses.

Effectively applying the slicing approach to one's own dataset thus minimally requires multiple slices within a batch and cluster, and at least part of one of these slices in another batch or cluster (Figure 2c; Figure S1). The slicing width can vary, for example, depending on the number of samples collected each year. Additionally, a low slicing angle is preferred since often there is a substantial number of confounding effects, and a lower slicing angle leads to slices crossing more batches and separation of confounding effects from variables of interest.

The benefit of the slicing approach over other strategies is that it allows convenient sequential analysis of the biobank and enables separation of variables of interest from confounding variables. However, the benefit of sequential analysis within the slicing approach disappears when samples from a single individual need to be aggregated within the same batch (e.g. Beirne et al., 2014; Nettle et al., 2015). We therefore determine whether aggregation of longitudinal samples from a single individual provides greater statistical power and precision in long-term studies to detect any within-individual, between-individual, year or cohort effects through simulations.

2.3 | Simulations

We used simulations run in R 3.3.1 (R Development Core Team, 2019) to determine the statistical power (i.e. ability to reject the null hypothesis when false) and precision (i.e. width of the distribution) to detect individual, year and cohort effects, using different sample allocation strategies (i.e. longitudinal samples aggregated in a single

batch, randomly allocated to batches, or 'sliced' across batches; see Data S1).

We simulated a population of 200 individuals in 10 cohorts that were sampled once a year for a maximum of 5 years, providing an equal sample size in all simulations. 'Telomere length' was used as an example response variable; however, this is applicable to any longitudinally measured continuous variable. Starting telomere length was drawn from a Gaussian distribution to fix between-individual standard deviation (*SD* = 1.00) and all individuals shared the same within-individual shortening rate of telomeres (0.06*1, scaled to *SD* = 1 parameter, =0.06 per year).

Year effects were simulated by taking 0.7 multiplied by a generated value drawn from a uniform distribution (between 0 and 1) for each year and added these to the response variable. In separate simulations, we replaced year with cohort effects (20 individuals per cohort) by taking 0.9 multiplied by a generated value from a uniform distribution (between 0 and 1) for each cohort. We chose to model 'year' and 'cohort' as possible biological confounds with experimentally induced variation. The choice to model such specific biology is rather arbitrary as we are simulating the confounding effect of 'batch of analysis' and biology. We also conduct additional simulations with varying strengths for year and cohort effects to determine the robustness of the results.

Individual probability of death was then modelled via telomere length associated with mortality (Equation 1) as:

$$y_i = \beta^{(\alpha^* x_i)}, \quad (1)$$

where *x* is the initial telomere length for *i*th individual, with a baseline probability of death (β) of 0.25 and a slope (α) of −0.23, providing mortality risk (y_i) per year. This resulted in the probability of death varying with ±2 *SD* telomere length from 0.14 to 0.36 per year. Death for each simulated individual was determined by drawing from a uniform distribution (ranging 0 to 1) to determine a simulated death. Mortality was partly determined by the response variable (to simulate selective disappearance from the population, determined by the between-individual age component, see next paragraph), with variable telomere lengths to start with (between-individual variation) and a set within-individual shortening (within-individual age component, see next paragraph).

We simulated the relationship between telomere length and age (in years) both within and between individuals. Between-individual effects were modelled using the mean age at which the individual's trait was measured, and within-individual effects as the age at which an individual's trait was measured minus the mean measurement age for that individual (van de Pol & Wright, 2009).

Simulations were run 5,000 times, for a varying number of samples (12, 24, 36, 48) per batch and simulated differences between batch means (batch attributable error, *SD*: 1, 2.5, 5, 10, 20, 40). This error is relatively high to ensure that we control for potential effects of batch attributable error when determining the variation in statistical power among sample allocation strategies. Simulations were repeated three times to obtain three separate results per sample allocation strategy.

The slicing strategy was simulated at an angle that resulted in at least three slices per batch. Note, to start the sample allocation, the first batch was filled by 3/4 with the first slice and by 1/4 with the second slice, where subsequent batches were filled by 1/4, 1/2 and 1/4 with subsequent slices (Figure 2c). Additional simulations were run with the slicing angle halved, slicing width halved, and a doubled sample size (*n* = 400).

The simulated data were analysed using linear mixed models in LME4 1.1–14 (Bates, Machler, Bolker, & Walker, 2015), where the model included random effects (at the intercept level) for individual (to control for repeated measurements on the same individual) and batch, and year or cohort was fitted as a fixed factor. Statistical power was determined by the number of significant values ($p < .05$) for each variable out of the total number of simulations (*n* = 5,000). It is important to understand the effect of sample allocation strategy on precision estimates, as well as statistical power. We therefore quantified precision as the width of the distribution of parameter estimates from the models run on the repeated simulated datasets, as the absolute difference between the 75% and 25% percentile divided by the median (note, a precision value closer to zero means higher precision).

Parameters of the simulations were manually optimized so that a statistical power of approximately 0.5 was achieved to detect between-individual effects for the random allocation strategy, determined by a *t*-value of less than −2 ($\alpha \approx 0.05$). This intermediate level of statistical power avoids thresholding effects at either end of the power spectrum (0 or 1). Such a simulation strategy maximizes the sensitivity in detecting any modulation in relative statistical power among sample allocation strategies, which is our focus rather than achieving a certain absolute statistical power.

3 | RESULTS

Our simulations tested the widely held assumption that aggregating longitudinal samples of the same individual in a single batch increases statistical power to detect within-individual effects (e.g. Herborn et al., 2014; Nettle et al., 2015). In simulations with year effects, the statistical power to detect within-individual effects was much lower when longitudinal samples were aggregated (mean statistical power ± *SD* across sample sizes and three runs per simulation = 0.059 ± 0.030) than when samples were sliced across batches (0.269 ± 0.008) or randomly allocated to batches (0.267 ± 0.007; Figure 3). For between-individual effects, again, the statistical power was much lower when longitudinal samples were aggregated in a single batch (0.138 ± 0.077) compared to when samples were sliced across batches (0.443 ± 0.007) or randomly allocated to batches (0.441 ± 0.007; Figure 3). The statistical power to detect year effects was higher when longitudinal samples were aggregated in a single batch (0.776 ± 0.008) or randomly allocated to batches (0.782 ± 0.014) than when sliced across batches (0.622 ± 0.012; Figure 3). However, a lower slicing angle (crossing four batches; 0.741 ± 0.009) and smaller slicing width (half a batch; 0.751 ± 0.007)

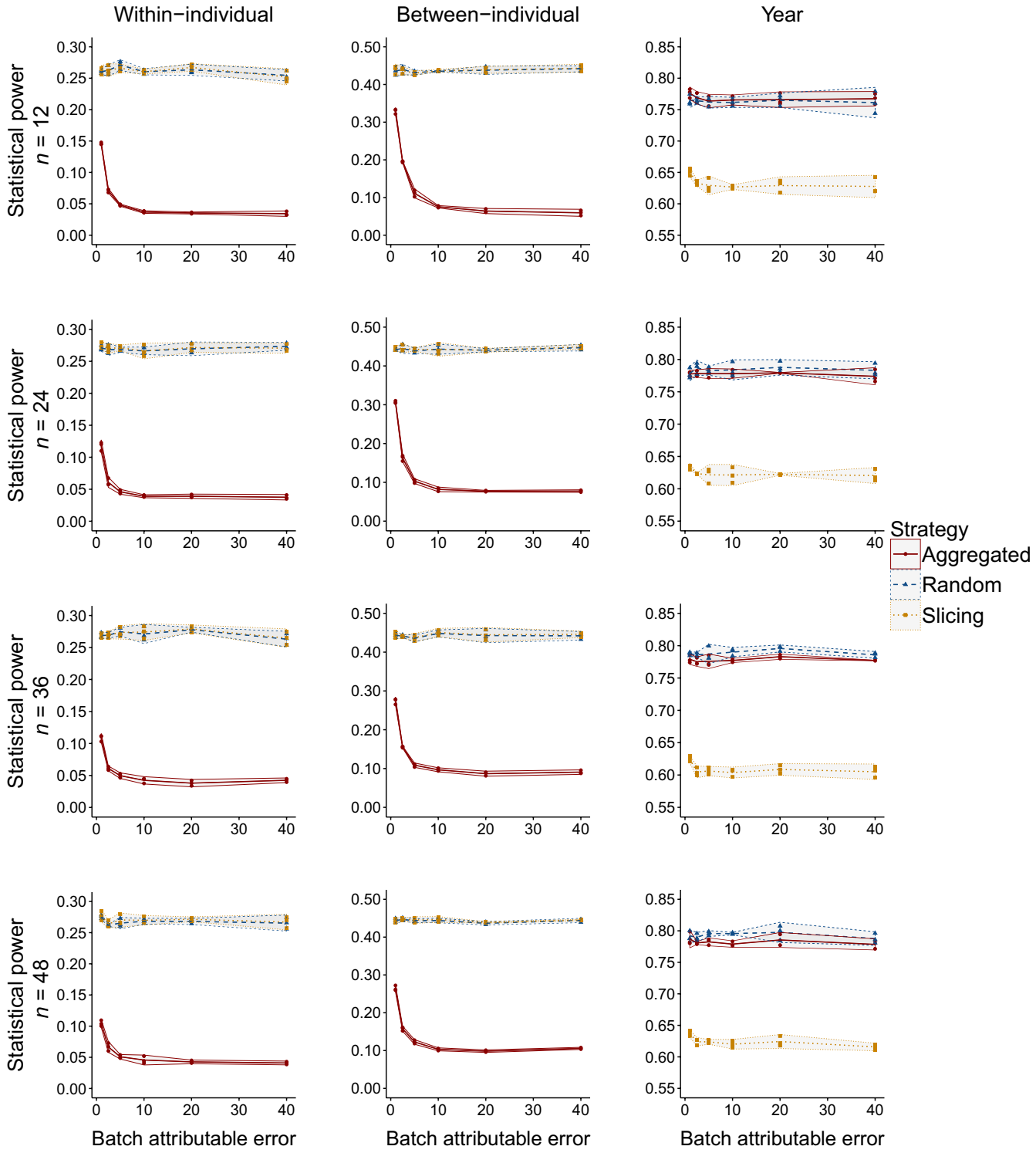

FIGURE 3 Statistical power analyses of simulated data for individual and year effects among four batch sizes (n = 12–48) using three sample allocation strategies: (1) aggregating samples per individual in the same batch (solid, red), (2) assigning samples randomly to batches (dashed, blue) or (3) slicing samples across batches with an angle that crosses two batches and a slicing width of a single batch (dotted, yellow). Raw data points from three separate simulations with mean statistical power per sample size are shown against among-batch variation, with 95% confidence intervals as shaded areas. Scales differ between year, within- and between-individual effects

resulted in a similar statistical power to detect year effects to aggregation of longitudinal samples and random allocation while maintaining statistical power to detect within- and between-individual effects (Figure 4).

In simulations with cohort effects, the statistical power to detect within- and between-individual effects was lower when slicing across batches (0.159 ± 0.032; 0.324 ± 0.020) compared to aggregation (0.557 ± 0.009; 0.390 ± 0.020) and randomization (0.542 ± 0.014;

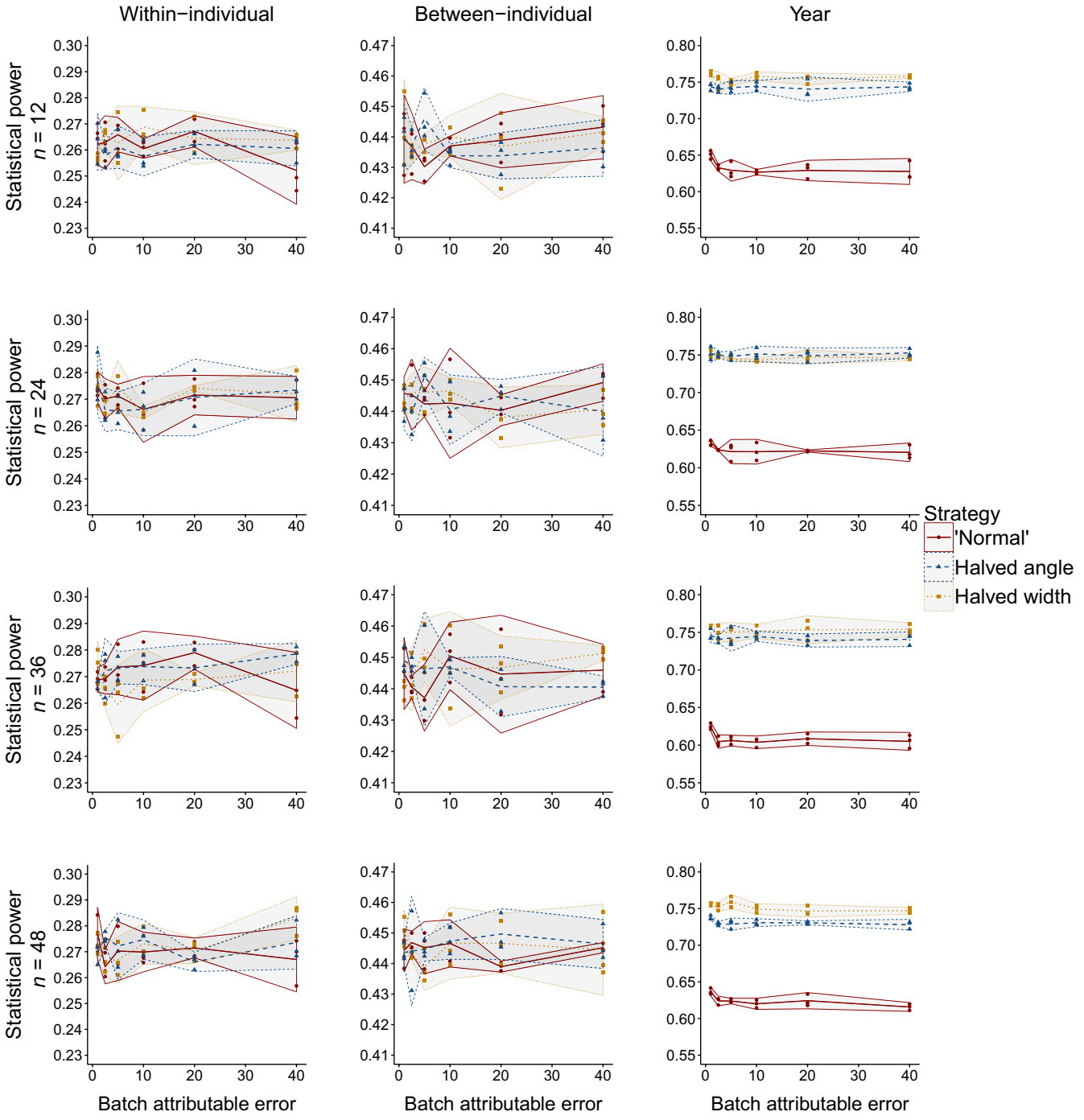

FIGURE 4 Statistical power analyses of simulated data for individual and year effects among four batch sizes (n = 12–48) using three different slicing parameters: (1) slicing angle that crosses two batches with a slicing width of a single batch (solid, red), (2) halved slicing angle which crosses four batches (dashed, blue) or (3) halved slicing width of half a batch (dotted, yellow). Raw data points from three separate simulations with mean statistical power per sample size are shown against among-batch variation, with 95% confidence intervals as shaded areas. Scales differ between year, within- and between-individual effects

0.423 ± 0.008) approaches (Figure S2). However, statistical power to detect cohort effects was greater for slicing (0.413 ± 0.009) and randomization (0.462 ± 0.008) compared to aggregation of longitudinal samples in a single batch (0.142 ± 0.044; Figure S2). A doubled sample size (n = 400), either for simulations with year or cohorts effects, increased statistical power but did not alter variation in statistical power among sample allocation strategies (Figures S3 and S4). Additionally, varying the strengths of year and cohort effects changed the statistical power, but not the variation among sample allocation strategies (Figures S5–S8).

In simulations with either year or cohort effects, the precision to estimate within- and between-individual effects followed

similar patterns to statistical power in the respective simulations, with greater precision for the approaches that showed greater statistical power (Figures S9 and S10). However, precision to estimate cohort and year effects was the opposite of statistical power, where approaches with lower statistical power showed greater precision to detect such effects (Figures S9 and S10). A doubled sample size (n = 400) increased precision but did not alter variation in precision among sample allocation strategies (Figures S11 and S12). Additionally, varying the strengths of year and cohort effects changed the precision, but not the variation among sample allocation strategies (Figures S13–S16).

Our slicing method performs similar to randomization of samples and outperforms aggregation of longitudinal samples to disentangle within- and between-individual effects when year effects apply, an objective shared by many longitudinal studies (Nussey, Froy, Lemaitre, Gaillard, & Austad, 2013; van de Pol & Wright, 2009). Simulations were run for a wide range of parameters and sample sizes (Figures 3 and 4; Figures S2–S16). When desirable, different parameter sets specific to current or future datasets can be included in the script provided (Data S1).

4 | DISCUSSION

The analysis of comprehensive long-term datasets is often complex due to cross-classified data structures and difficulties in separating variation in sources of interest from confounding variables, such as separating year from batch effects. Our simulations clearly demonstrate that statistical power was greater for within- and between-individual effects when samples were randomized or sliced across batches when year effects apply. The reduction in statistical power for aggregation of longitudinal samples in a single batch to detect such effects can be explained by the confounding of the year and batch effect. Conversely, in simulations with cohort effects, there was greater statistical power for within- and between-individual effects when samples were aggregated within a single batch or randomized, compared to slicing across batches. The lower statistical power for slicing to detect such effects is the consequence of a low number of cohorts in our simulations, where cohorts are sequentially ordered instead of mixed among slices, which results in confounding effects between cohort and batch within the slicing approach. A higher number of cohorts in the simulations will lead to a mixture of cohorts among slices and result in similar statistical power to detect within- and between-individual effects for all three sample structuring strategies. This highlights the importance choosing appropriate slicing angles and widths, ensuring adequate variation of potential confounds (e.g. cohort/year) in a single batch.

The greater statistical power to detect within-individual effects for slicing and randomization when year effects apply was the consequence of appropriate statistical methodology, accounting for batch, individual and year through fixed and random effects. These results disprove the assumption that samples from a single individual need to be analysed in the same batch for greater statistical power to detect within-individual effects (e.g. Beirne et al., 2014; Nettle et al., 2015). Such efforts will reduce the statistical power of the study and generate unnecessary effort in picking specific samples, which increases the likelihood of technical errors (e.g. sample mix-ups, freeze/thawing effects, transcription or pipetting errors).

The statistical power to detect year effects was greater when samples were aggregated in a single batch or randomized across batches compared to slicing. However, when the slicing angle and width decreased, there was no difference in statistical power to detect year effects compared with aggregation and randomization approaches. This is because a lower slicing angle and smaller width reduces the confound between batch and year, as a slice crosses more batches or there are more slices per batch. In contrast, the notion that longitudinal samples should not be aggregated in the same batch becomes particularly pronounced when cohort effects occur. The effort of grouping samples from a single individual together collects cohorts together (an individual's cohort is fixed) in a batch thus reducing the statistical power to distinguish between different cohorts, even though this increases precision. Random allocation of samples and slicing have a substantially greater statistical power to detect cohort effects due to a higher mixture of cohorts within the same batch. For telomere biology especially, estimating cohort effects reliably is important as it can affect telomere length strongly (Spurgin et al., 2017), but cohort effects are not always estimated. All these results are robust against a variety of batch errors, sample sizes and strengths of year and cohort effects.

4.1 | Integral approach to growing biobank analysis

The optimal sample structuring strategy for analysing long-term datasets depends on the structure of the data and questions of interest. However, in the majority of long-term datasets, slicing has benefits over other structuring strategies by overcoming problems with confounding variables and cross-classified data structures which commonly occur in the analysis of long-term studies.

The assumption that longitudinal samples should be aggregated in a single batch could hinder the slicing approach, but our simulations have disproven this assumption. Slicing performs, in terms of statistical power and precision, equally well to randomization when applying correct slicing parameters (i.e. low width and angle). Slicing across batches and clusters and bridging them together provides the slicing approach with statistical power to disentangle confounding effects.

The key benefit of slicing over randomization is that slicing allows separate analysis of current data and flexible inclusion of these data into future analyses without completely confounding the analysis. Furthermore, slicing allows sequential analysis of samples, which only need to be analysed once, preventing complicated sample labelling and placing among clusters, reducing sample volume required and avoiding any defrosting issues and therefore reducing the potential for human error.

Slicing has some potential limitations. For example, substantial differences among years in the number of samples collected could limit the ease with which the slicing approach is applied. Additionally, a failed analysis of samples (e.g. plate failure leading to sample loss during analysis) using slicing results in missing data within a certain time window, whereas with randomization this is scattered across the dataset. While slicing performs similarly to randomization in terms of statistical power and precision, we think that slicing is more practical with merits (i.e. sequential analysis, statistical independence) that outweigh the limitations. We stipulate that, because of sequential analysis in our slicing approach, hypotheses need to be pre-defined and power analyses conducted before experimental and statistical analysis (Fraser, Parker, Nakagawa, Barnett, & Fidler, 2018 and references therein).

The use of mixed models is common in the analysis of longitudinal datasets, especially in ecology (Bolker et al., 2009; Gelman & Hill, 2006). We highlight the use of mixed models because they are necessary when using the slicing approach to account adequately for experimental and environmental variation. The combination of slicing and mixed models in long-term studies allows analysis of commonly occurring cross-classified data structures that arise due to hierarchical biology mixed with cross-classified data collection and analysis. Interpretation of the variance components in these models depends on a crossed or nested design (Schielzeth & Nakagawa, 2013), where the random effect structure can be used to account for potentially confounding experimental and environmental variables with cluster effects (e.g. storage duration, batch). The failure to include these effects can inflate type I and type II errors when there is a temporal, spatial or other spurious correlation with any independent variable.

5 | CONCLUSIONS

A major current challenge in long-term studies is analysing data as it is collected while also including it in future analyses without creating uncontrollable variation, allowing comparison of results over multiple years or even decades. This requires the ability to compare differentially timed analyses that are potentially biased by confounding cluster effects. Our study shows the importance of considering the structure of samples among clusters and batches in long-term studies. Our slicing approach retains statistical independence and accounts for among-cluster variation in the sequential analysis of growing biobanks. Slicing also provides similar statistical power and precision to detect cohort, year, within- and between-individual effects to randomization, if analysed using appropriate statistical mixed models and consistent methodology to control for confounding effects. A single sample's scientific value increases through this approach, as it can be used separately in current studies, but can also be included in subsequent studies, providing sustainable (re-) use of collected data. The approach we propose here (slicing and mixed models) is easy to apply and improves the potential for these growing biobanks to address important ecological and evolutionary questions.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge feedback on an earlier version of the manuscript from Dan Nussey. We also thank two anonymous reviewers for comments which greatly improved the manuscript. This work was supported by a Leeds Anniversary Research Scholarship to S.H.J.v.L., a Sir Henry Wellcome (WT107400MA) Fellowship and a University of Sheffield Vice-Chancellor's Fellowship to M.J.P.S., a NERC grant to H.L.D. and T.B. (NE/P011284/1), a NERC grant (NE/J024597/1) to T.B. and J.S., and a NERC grant (NE/N013832/1) to T.B. and M.J.P.S. The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

Conception/design: S.H.J.v.L., J.S., T.B., M.J.P.S. and H.L.D. Methodology: S.H.J.v.L., M.J.P.S. Analysis and interpretation: S.H.J.v.L., M.J.P.S. and H.L.D. Drafting/Revising paper: S.H.J.v.L., H.F., J.S., T.B., M.J.P.S. and H.L.D. All authors contributed critically to the manuscript and gave final approval for publication.

DATA AVAILABILITY STATEMENT

Supporting Table S1 and Figures S1–S16 are provided in the supporting information. Simulation R scripts (Data S1) are provided in the supporting information, archived at GitHub (https://github.com/DugdaleResearchGroup/Slicing), and available at Zenodo (https://doi.org/10.5281/zenodo.3606156; van Lieshout et al., 2020).

ORCID

Sil H. J. van Lieshout https://orcid.org/0000-0003-4136-265X
Hannah Froy https://orcid.org/0000-0003-2965-3526
Julia Schroeder https://orcid.org/0000-0002-4136-843X
Terry Burke https://orcid.org/0000-0003-3848-1244
Mirre J. P. Simons https://orcid.org/0000-0001-7406-7708
Hannah L. Dugdale https://orcid.org/0000-0001-8769-0099

REFERENCES

Andraud, M., Casas, M., Pavio, N., & Rose, N. (2014). Early-life hepatitis E infection in pigs: The importance of maternally-derived antibodies. *PLoS ONE*, *9*, e105527. https://doi.org/10.1371/journal.pone.0105527

Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Beirne, C., Delahay, R., Hares, M., & Young, A. (2014). Age-related declines and disease-associated variation in immune cell telomere length in a wild mammal. *PLoS ONE*, *9*, e108964. https://doi.org/10.1371/journal.pone.0108964

Berry, O., England, P., Marriott, R. J., Burridge, C. P., & Newman, S. J. (2012). Understanding age-specific dispersal in fishes through hydrodynamic modelling, genetic simulations and microsatellite DNA analysis. *Molecular Ecology*, *21*, 2145–2159. https://doi.org/10.1111/j.1365-294X.2012.05520.x

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*, 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Boonekamp, J. J., Mulder, G. A., Salomons, H. M., Dijkstra, C., & Verhulst, S. (2014). Nestling telomere shortening, but not telomere length, reflects developmental stress and predicts survival in wild birds. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20133287. https://doi.org/10.1098/rspb.20133287

Clutton-Brock, T., & Sheldon, B. C. (2010). Individuals and populations: The role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, *25*, 562–573. https://doi.org/10.1016/j.tree.2010.08.002

Dantzer, B., Newman, A. E. M., Boonstra, R., Palme, R., Boutin, S., Humphries, M. M., & McAdam, A. G. (2013). Density triggers maternal hormones that increase adaptive offspring growth in a wild mammal. *Science*, *340*, 1215–1217. https://doi.org/10.1126/science.1235765

Fairlie, J., Holland, R., Pilkington, J. G., Pemberton, J. M., Harrington, L., & Nussey, D. H. (2016). Lifelong leukocyte telomere dynamics and survival in a free-living mammal. *Aging Cell*, *15*, 140–148. https://doi.org/10.1111/acel.12417

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLoS ONE*, *13*, e0200303. https://doi.org/10.1371/journal.pone.0200303

Gelman, A. (2005). Analysis of variance – Why it is more important than ever. *Annals of Statistics*, *33*, 1–31. https://doi.org/10.1214/009053604000001048

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York, Cambridge: Cambridge University Press.

Greenland, S., Robins, J. M., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, *14*, 29–46. https://doi.org/10.1214/ss/1009211805

Herborn, K. A., Heidinger, B. J., Boner, W., Noguera, J. C., Adam, A., Daunt, F., & Monaghan, P. (2014). Stress exposure in early post-natal life reduces telomere length: An experimental demonstration in a long-lived seabird. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20133151. https://doi.org/10.1098/rspb.2013.3151

Martin, J. G. A., Nussey, D. H., Wilson, A. J., & Reale, D. (2011). Measuring individual differences in reaction norms in field and experimental studies: A power analysis of random regression models. *Methods in Ecology and Evolution*, *2*, 362–374. https://doi.org/10.1111/j.2041-210X.2010.00084.x

Nettle, D., Monaghan, P., Gillespie, R., Brilot, B., Bedford, T., & Bateson, M. (2015). An experimental demonstration that early-life competitive disadvantage accelerates telomere loss. *Proceedings of the Royal Society B: Biological Sciences*, *282*, 20141610. https://doi.org/10.1098/rspb.2014.1610

Nussey, D. H., Froy, H., Lemaitre, J. F., Gaillard, J. M., & Austad, S. N. (2013). Senescence in natural populations of animals: Widespread evidence and its implications for bio-gerontology. *Ageing Research Reviews*, *12*, 214–225. https://doi.org/10.1016/j.arr.2012.07.004

Pauliny, A., Devlin, R. H., Johnsson, J. I., & Blomqvist, D. (2015). Rapid growth accelerates telomere attrition in a transgenic fish. *BMC Evolutionary Biology*, *15*, 159. https://doi.org/10.1186/s12862-015-0436-8

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge, UK: Cambridge University Press.

R Development Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rius-Ottenheim, N., Houben, J. M. J., Kromhout, D., Kafatos, A., van der Mast, R. C., Zitman, F. G., … Giltay, E. J. (2012). Telomere length and mental well-being in elderly men from the Netherlands and Greece. *Behavior Genetics*, *42*, 278–286. https://doi.org/10.1007/s10519-011-9498-6

Schielzeth, H., & Nakagawa, S. (2013). Nested by design: Model fitting and interpretation in a mixed model era. *Methods in Ecology and Evolution*, *4*, 14–24. https://doi.org/10.1111/j.2041-210x.2012.00251.x

Schroeder, J., Burke, T., Mannarelli, M. E., Dawson, D. A., & Nakagawa, S. (2012). Maternal effects and heritability of annual productivity. *Journal of Evolutionary Biology*, *25*, 149–156. https://doi.org/10.1111/j.1420-9101.2011.02412.x

Schroeder, J., Nakagawa, S., Rees, M., Mannarelli, M. E., & Burke, T. (2015). Reduced fitness in progeny from old parents in a natural population. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 4021–4025. https://doi.org/10.1073/pnas.1422715112

Simons, M. J. P., Winney, I., Nakagawa, S., Burke, T., & Schroeder, J. (2015). Limited catching bias in a wild population of birds with near-complete census information. *Ecology and Evolution*, *5*, 3500–3506. https://doi.org/10.1002/ece3.1623

Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: SAGE.

Spurgin, L. G., Bebbington, K., Fairfield, E. A., Hammers, M., Komdeur, J., Burke, T., … Richardson, D. S. (2017). Spatio-temporal variation in lifelong telomere dynamics in a long-term ecological study. *Journal of Animal Ecology*, *87*, 187–198. https://doi.org/10.1111/1365-2656.12741

Sudyka, J., Arct, A., Drobniak, S., Dubiec, A., Gustafsson, L., & Cichon, M. (2014). Experimentally increased reproductive effort alters telomere length in the blue tit (*Cyanistes caeruleus*). *Journal of Evolutionary Biology*, *27*, 2258–2264. https://doi.org/10.1111/jeb.12479

Swanson, C. R., Li, K., Unger, T. L., Gallagher, M. D., Van Deerlin, V. M., Agarwal, P., … Chen-Plotkin, A. S. (2015). Lower plasma apolipoprotein A1 levels are found in Parkinson's disease and associate with apolipoprotein A1 genotype. *Movement Disorders*, *30*, 805–812. https://doi.org/10.1002/mds.26022

Takizawa, T., Mitsumori, K., Takagi, H., Nasu, M., Yasuhara, K., Onodera, H., … Hirose, M. (2004). Sequential analysis of testicular lesions and serum hormone levels in rats treated with a *Psoralea corylifolia* extract. *Food and Chemical Toxicology*, *42*, 1–7. https://doi.org/10.1016/S0278-6915(03)00220-5

Telfer, S., Birtles, R., Bennett, M., Lambin, X., Paterson, S., & Begon, M. (2008). Parasite interactions in natural populations: Insights from longitudinal data. *Parasitology*, *135*, 767–781. https://doi.org/10.1017/S0031182008000395

Tollenaere, C., Ivanova, S., Duplantier, J. M., Loiseau, A., Rahalison, L., Rahelinirina, S., & Brouat, C. (2012). Contrasted patterns of selection on MHC-linked microsatellites in natural populations of the Malagasy plague reservoir. *PLoS ONE*, *7*, e32814. https://doi.org/10.1371/journal.pone.0032814

Underwood, A. J. (1997). *Experiments in ecology: Their logical design and interpretation using analysis of variance*. Cambridge, UK: Cambridge University Press.

van de Pol, M. (2012). Quantifying individual variation in reaction norms: How study design affects the accuracy, precision and power of random regression models. *Methods in Ecology and Evolution*, *3*, 268–280. https://doi.org/10.1111/j.2041-210X.2011.00160.x

van de Pol, M., & Wright, J. (2009). A simple method for distinguishing within- versus between-subject effects using mixed models. *Animal Behaviour*, *77*, 753–758. https://doi.org/10.1016/j.anbehav.2008.11.006

van Lieshout, S. H. J., Froy, H., Schroeder, J., Burke, T., Simons, M. J. P., & Dugdale, H. L. (2020). Data from: Slicing: A sustainable approach to structuring samples for analysis in long-term studies. *Zenodo*. https://doi.org/10.5281/zenodo.3606156

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*, 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: van Lieshout SHJ, Froy H, Schroeder J, Burke T, Simons MJP, Dugdale HL. Slicing: A sustainable approach to structuring samples for analysis in long-term studies. *Methods Ecol Evol*. 2020;11:418–430. https://doi.org/10.1111/2041-210X.13352